

What is claimed is:

1. A handwritten character recognition network for inferring parts of handwriting from visual observations with a common hidden variable having plural discrete states, comprising:

at least one mixture of Bayesian networks (MBN) that encodes probabilities of observing the visual observations corresponding to a handwritten character;

the MBN comprising:

a plurality of hypothesis-specific Bayesian networks (HSBNs), each of the HSBNs encoding probabilities of observing the visual observations corresponding to a handwritten character and given the common hidden variable being in a respective one of its states;

an aggregator that combines outputs of the HSBNs to produce an MBN output of the MBN;

each one of the HSBNs comprises:

plural nodes, each of the plural nodes corresponding to an associated visual observation element, and

at least some of the plural nodes having dependencies with others of the plural nodes within the one HSBN,

an aggregator connected to outputs of the nodes of the one HSBN to provide the output of the one HSBN, the nodes having inputs that receive the state of a respective one of the visual observation elements of a current one of the visual observations.

2. The handwriting recognition network claim 1, at least one node of the plural nodes storing probability parameters that correspond to a relationship between the visual observation element of the at least one node and the visual observation element of at least one other of the nodes in the respective HSBN.

3. The handwriting recognition network of claim 1, at least one node of the plural nodes of a respective HSBN storing probability parameters that define how a respective state of the associated visual observation element depends on the respective handwritten character for the respective MBN.
4. The handwriting recognition network of claim 3, the visual observation elements further comprising a number of start-stop features associated with the beginning and end of at least one character stroke for the respective character.
5. The handwriting recognition network of claim 4, the number of start-stop features comprising $4*k$, where k is the number of character strokes for the respective character, the common hidden variable having a fixed predetermined number of the states.
6. The handwriting recognition network of claim 5, the visual observation elements further comprising at least one mid-point features associated with part of the character stroke between the beginning and end of at least some of the character strokes.
7. The handwriting recognition network of claim 4, the visual observation elements further comprising curvature features associated with part of the at least some of the character strokes, the common hidden variable having a fixed predetermined number of states.
8. The handwriting recognition network of claim 4, the number of states for the common hidden variable being functionally related to the number of training examples for the respective character.

9. The handwriting recognition network of claim 8, the number of states for the common hidden variable being 2 for 1 to 300 training examples, 3 for 301 to 1000 training examples, 4 for 1001 to 3000 training examples, 5 for 3001 to 5000 training examples, and the number of training examples divided by one-thousand for more than 5000 training examples, with a maximum of 10 states.

10. The handwriting recognition network of claim 3, the probability parameters in different ones of the HSBNs differ to reflect different states of the common external hidden variable represented by the different ones of the HSBNs.

11. The handwriting recognition network of claim 1, the one HSBN having a restricted structure in which no node associated with a visual observation element for a character stroke is dependent on another node associated with a visual observation element of an earlier character stroke of the respective handwritten character.

12. The handwriting recognition network of claim 1, further comprising a system for performing character completion according to a set of MBN arrays selected as a function of the number character strokes in an input handwritten character of the visual observation.

13. The handwriting recognition network of claim 12, the system for performing character completion restricting each of the HSBNs so that no node associated with a visual observation element for a character stroke is dependent on another node associated with a visual observation element of an earlier character stroke of the input handwritten character.

14. The handwriting recognition network of claim 1, the common hidden variable being external in that the common hidden variable is not represented by any of the nodes in the mixture of Bayesian networks.

15. The handwriting recognition network of claim 1, each HSBN being associated with an HSBN score, and:

each of the HSBNs further comprises an inference input defining observed data corresponding to the visual observations and an inference output corresponding to the likelihood of an visual observation corresponding to a handwritten character and given the common hidden variable being in one of its states corresponding to the HSBN; and

the mixture of Bayesian networks further comprises a weight multiplier which weights the inference output of each HSBN by a corresponding HSBN score and combines the weighted HSBN inference outputs into a single inference output of the mixture of Bayesian networks.

16. The handwriting recognition network of claim 15, the HSBN score corresponding to the likelihood of the common hidden variable being in the corresponding one of the states of the common hidden variable.

17. The handwriting recognition network of claim 16, the HSBN score reflects the goodness of the corresponding HSBN at predicting observed data representing states of the observed variables.

18. The handwriting recognition network of claim 17, the HSBN score is computed by the mixture of Bayesian networks.

19. The handwriting recognition network of claim 1, the number of the HSBNs in the MBN is selected to optimize the goodness of the mixture of Bayesian network at predicting observed data representing states of the observed variables.

20. The handwriting recognition network of claim 19, the number of HSBNs in the MBN corresponds to the number of states of the common hidden variable for the respective MBN.

21. The handwriting recognition network of claim 1, the plurality of nodes of different ones of the HSBNs represent the same set of hidden and observed variables.

22. A handwriting recognition system, comprising:

means for encoding probabilities of observing sets of visual observations for a predetermined character;

the means for encoding comprising plural means for modeling a hypothesis that a common hidden variable corresponding to a handwritten character associated with the means for encoding is in a respective one of a plurality of discrete states;

each of the means for modeling comprising plural means for storing probability parameters that define relationships between visual features of the handwritten character;

at least some of the means for storing having dependencies with others of the means for storing in each respective means for modeling;

each means for modeling further comprising means for aggregating outputs from the plural means for storing and providing outputs of each respective means for modeling; and

means for aggregating the outputs of the respective means for modeling to produce a corresponding output of each respective means for encoding indicative of the probability that an input visual observation corresponds to the handwritten character defined by the common hidden variable .

23. A system for training a mixture of Bayesian networks (MBN), comprising:

the mixture of Bayesian network comprising a plurality of hypothesis-specific Bayesian networks (HSBNs), each of the HSBNs modeling a hypothesis that a common

hidden variable corresponding to a handwritten character associated with the MBN is in a respective one of a plurality of discrete states;

each of the HSBNs comprising a plurality of nodes that correspond to visual features of the handwritten character, each one of the plurality of nodes in a respective HSBN storing probability parameters that indicate how a visual feature of the one node depends on the handwritten character;

a parameter search component that identifies a set of changes in the probability parameters that improve the goodness of each of the HSBNs in predicting visual observations of the handwritten character;

a parameter modification component that modifies the probability parameters based on the identified set of changes;

a scoring system that computes a structure score for each HSBN in the MBN that reflects the goodness of the each respective HSBN in predicting visual observations according to a structure of each HSBN;

a network adjuster that searches for changes in dependencies between nodes of each HSBN that improve the structure score and modifies the dependencies so as to improve the structure score.

24. The system of claim 23, the parameter modification component and the network adjuster cooperating for each HSBN so as to interleave the search for changes in the probability parameters and the changes in the dependencies among the nodes.

25. A method of training a mixture of Bayesian networks (MBN) to facilitate recognition of handwritten characters, the MBN encoding probabilities of observing the sets of visual observations corresponding to a handwritten character and comprising a plurality of hypothesis-specific Bayesian networks (HSBNs), each HSBN including a plurality of nodes having probability parameters with dependencies between at least some of the nodes to model a hypothesis that a common hidden variable corresponding

to a handwritten character is in a respective one of a plurality of discrete states, the method comprising:

for each one of the HSBNs:

conducting a parameter search for a set of changes in the probability parameters which improves the goodness of the one HSBN in predicting the visual observations, and

modifying the probability parameters of the one HSBN accordingly; and

for each one of the HSBNs:

computing a structure score of the one HSBN reflecting the goodness of the one HSBN in predicting the visual observations,

conducting a structure search for a change in the causal links which improves the structure search score, and

modifying the causal links of the one HSBN based on the structure search.

26. The method of claim 25, the computing a structure score of the one HSBN further comprises:

computing expected complete model sufficient statistics (ECMSS) based on the visual observations;

computing sufficient statistics for the one HSBN based on the ECMSS; and

computing the structure score based on the sufficient statistics.

27. The method of claim 26, the plurality of nodes in the one HSBN further comprising discrete hidden and observed variables having respective states, the computing the ECMSS further comprising:

computing the probability of each combination of states of discrete hidden and observed variables of the nodes of the one HSBN;

forming a vector for each observed case in the set of visual observations, each entry in the vector corresponding to a particular one of the combinations of the states of the discrete variables; and

summing the vectors over plural cases of the visual observations.

28. The method of claim 27, at least some of the plurality of nodes in the one HSBN further comprising continuous variables, each entry in the vector is formed to have plural sub-entries comprising:

(a) the probability of the one combination of the states of the discrete variables,

(b) sub-entry vectors representing the states of the continuous variables.

29. The method of claim 28, further comprising computing by inference in the MBNs the probability of the one combination of the states of the discrete variables.

30. The method of claim 28, wherein each of the plural sub-entries is formed such that the sub-entry vector has a vector multiplier corresponding to the probability of the one combination of the states of the discrete variables.

31. The method of claim 30, the computing sufficient statistics based on the ECMSS comprises computing from the ECMSS at least one of the following:

- (a) mean,
- (b) scatter,
- (c) sample size.

32. The method of claim 25, the conducting a parameter search and the modifying the probability parameters are repeated consecutively until a parameter search convergence criteria is met.

33. The method of claim 32, further comprising:

repeating the conducting a parameter search, the computing the structure score and the conducting a structure search until a structure search convergence criteria is met.

34. The method of claim 33, the parameter search convergence criteria is a determination of whether the parameter search has converged at a local optimum.

35. The method of claim 33, the parameter search convergence criteria is a determination of whether the parameter search has been repeated a number of times.

36. The method of claim 35, the number of times is a set number.

37. The method of claim 35, the number of times is a function of the number of times the structure search has been repeated for the one HSBN.

38. The method of claim 35, the parameter search convergence criteria limits the repetition of the parameter search to a limited number of repetitions and the parameter search is repeated after convergence of the structure search.

39. The method of claim 33, the structure search convergence criteria comprises a determination of whether the structure score has worsened since a prior repetition of the structure search step.

40. The method of claim 33, the structure search criteria comprises a determination of whether a current performance of the structure search has changed any of the dependencies in the one HSBN.

41. The method of claim 25, further comprising:
repeating the conducting a parameter search, the computing the structure score and the conducting a structure search until a structure search convergence criteria is met.
42. The method of claim 25, the conducting a structure search further comprises:
attempting different modifications of the dependencies at each node of the one HSBN;
for each one of the different modifications, computing the structure score of the one HSBN; and
saving those modifications providing improvements to the structure score.
43. The method of claim 25, further comprising computing a combined score of the MBN from the structure scores of computed for each of the plurality of HSBNs.
44. The method of claim 43, further comprising choosing a different number of states of the discrete hidden and observed variables and repeating the parameter and the structure search steps to generate a different MBN and scores thereof for the different numbers of states of the discrete variables.
45. The method of claim 44, further comprising choosing the MBN having the highest score.
46. The method of claim 44, further comprising weighting inference outputs of the different mixtures of Bayesian networks in accordance with their individual scores.
47. The method of claim 35, further comprising repeating the parameter search when conducting a structure search results in a change in the structure of the one HSBN.

48. The method of claim 47, the conducting of each of the parameter search and the structure searching being repeated either a fixed number of times or until an associated convergence criteria is met.

49. The method of claim 47, the parameter search is repeated by a number of times functionally related to the number of times the structure search has been repeated.

50. The method of claim 25, further comprising repeating the conducting of the parameter search and the conducting of the structure search and interleaving repetitions of the parameter search and the structure search.

51. The method of claim 25, further comprising initializing each of the HSBNs by
(a) defining a causal link from each node corresponding to a hidden variable to each node corresponding to a continuous observed variable; and
(b) initializing the probability parameters in each of the nodes.

52. The method of claim 51, each of the plurality of nodes in the one HSBN corresponding to a character completion characteristic of the handwritten character, the method further comprising enforcing a restriction that no node is dependent on another node in the one HSBN that is associated with a character completion characteristic of an earlier character stroke of the handwritten character.

53. The method of claim 25 wherein the step of performing the parameter search comprises searching for a change in the probability parameters in each node which improves the performance of the one HSBN in predicting the visual observations.

54. The method of claim 25, the common hidden variable is a common external discrete hidden variable not represented by any node in the MBN, the number of HSBNs

in the MBN is equal to the number of states of the common external discrete hidden variable.

55. The method of claim 25, further comprising, for each HSBN, determining an optimum number m of HSBNs in the MBN, whereby m can be different for each MBN.

56. A computer-readable medium storing computer-executable instructions for performing the method of claim 25.

57. A system for inferring a handwritten character from visual observations, comprising:

mixtures of Bayesian networks (MBNs), each MBN encoding the probabilities of the visual observations associated with a handwritten character, each of the MBNs associated with a common hidden variable;

each of the MBNs comprising a plurality of hypothesis-specific Bayesian networks (HSBNs) that model a hypothesis that the common hidden variable corresponding to a handwritten character is in a respective one of a plurality of discrete states.

58. The system of claim 57, each of the HSBNs further comprising a plurality of nodes, each of the nodes in a respective one of the HSBNs corresponding to an associated visual observation element of the handwritten character, at least some of the nodes in a respective one of the HSBNs having dependencies with others of the nodes in the one HSBN.

59. The system of claim 58, at least one node of the one HSBN storing probability parameters that define how a respective state of the associated visual observation element depends on the handwritten character.

Docket No. MS30475.8

60. The system of claim 58, the visual observation elements further comprising at least one character completion characteristics associated with at least one character stroke of the handwritten character.

61 The system of claim 58, the common hidden variable is a common external hidden variable not included in any of the HSBNs, the number of HSBNs in the MBN is equal to the number of states of the common hidden variable.